DOCUMENT RESUME

ED 045 701 TM 000 267

AUTHOR Knapp, Thomas R.; Schafer, William D.
TITLE Two New Formulas for F in Analysis of Covariance for Two Groups and One Covariable.
PUB DATE Feb 71
NOTE 6p.; Paper presented at the annual meeting of the American Educational Research Association, New York, February 1971

EDRS PRICE EDRS Price MF-$0.25 HC-$0.40
DESCRIPTORS *Analysis of Covariance, Analysis of Variance, Correlation, Multiple Regression Analysis, Post Testing, Pretests, Research Design, *Research Methodology, *Statistical Analysis, *Tests of Significance

ABSTRACT

Two theorems concerning F in analysis of covariance with two groups (experimental and control) and one covariable (pretest score) are presented. The first shows explicitly that F is a direct function of the ratio of the variance about the regression line for the total sample to the variance about the within-group regression line. The second demonstrates the relationship between F and the corresponding F, which would result from analysis of variance using the appropriate pretest-posttest gain scores in the same situation. (Author)

TWO NEW FORMULAS FOR F IN ANALYSIS OF COVARIANCE FOR TWO GROUPS AND ONE COVARIABLE

Thomas R. Knapp
University of Rochester

and

William D. Schafer
University of Maryland

One of the most frequently used experimental designs in educational research is the pretest-posttest control group design (with or without random assignment of individuals to the treatment groups). The common way to analyze the data for such a design is to carry out a t-test of the significance of the difference between the mean "gain" for the experimental group and the mean "gain" for the control group. But as Campbell and Stanley (1963) and others have pointed out, the better procedure is to use analysis of covariance with the posttest score as the dependent variable and the pretest score as the covariable. Gain score analysis is less precise (Feldt, 1958), incorrectly assumes that the within-group regression coefficient is equal to one (Edwards, 1968), and is limited to applications in which the pretest is the same test as the posttest or an equivalent form thereof.

The following theorems concerning F in analysis of covariance for two groups and one covariable should be of interest to researchers. The first shows explicitly that the covariance F is a direct function of the ratio of the variance about the regression line for the total sample to the variance about the within-group regression line. The second demonstrates the mathematical relationship between gain score F and covariance F.

Theorem 1 Let N_1 and N_2 be the number of individuals in group 1 (the experimental group) and group 2 (the control group), respectively. Let $s_{Y_T}^2$ be the variance of the posttest scores for the total group of $N_1 + N_2$ ($=N$) individuals, and let $s_{Y_W}^2$ be the within group variance of the posttest scores. Let r_{XY_T} and r_{XY_W} be the Pearson product-moment correlation coefficients between the pretest (X) and the posttest (Y) for total group and within-group, respectively. Then, under the usual assumptions for the analysis of covariance,

$$F_{1,N-3} = (N-3)\left[\frac{s_{Y_T}^2 \; (1-r_{XY_T}^2)}{s_{Y_W}^2 \; (1-r_{XY_W}^2)} - 1\right] \qquad (1)$$

<u>Proof</u>

$$F = \frac{\frac{ABSSY}{df_B}}{\frac{AWSSY}{df_W}}$$

, where ABSSY and AWSSY are the adjusted sums of squares for between and within groups, respectively, and df_B and df_W are the corresponding numbers of degrees of freedom

$$F_{1,N-3} = (N-3)\ \frac{ABSSY}{AWSSY}$$

$$= (N-3)\ \frac{ATSSY - AWSSY}{AWSSY}$$

, where ATSSY is the adjusted total sum of squares

$$= (N-3)\left[\frac{ATSSY}{AWSSY} - 1\right]$$

$$= (N-3)\left[\frac{Ns_{Y\cdot X_T}^2}{Ns_{Y\cdot X_W}^2} - 1\right].$$

(McNemar, 1962); where $s_{Y\cdot X_T}^2$ and $s_{Y\cdot X_W}^2$ are the variances about the regression lines for total and within, respectively

$$= (N-3)\left[\frac{s_{Y_T}^2 \; (1-r_{XY_T}^2)}{s_{Y_W}^2 \; (1-r_{XY_W}^2)} - 1\right]$$

Formula (1) can be used by the researcher to obtain the covariance F directly as a function of N , $s^2_{Y_T}$, $s^2_{Y_W}$, r_{XY_T} , and r_{XY_W} , all of which are relatively easy to calculate. For real data the within-group posttest variances and the within-group correlations will not be exactly the same. Tests of homogeneity of variance about the within-group regression line and homogeneity of the within-group regression coefficients should be carried out (Wilson and Carry, 1969). If the homogeneity assumptions are satisfied, $s^2_{Y_W}$ and r_{XY_W} can be determined by the "Model A" formulas given by Olkin (1967) for pooling variances and correlations. If these assumptions are not satisfied, the analysis of covariance technique should not be used.

__Theorem 2__ Let G_1 and G_2 be the "gain" scores, Y-X, for the members of group 1 and group 2, respectively. Let $F_{1,N-2}$ be the gain score $\underline{F}$ and let $F_{1,N-3}$ be the covariance $\underline{F}$ for the same data. Then, if $s^2_{X_T}$ and $s^2_{X_W}$ are the variances of the pretest scores for total and within groups, respectively,

$$F_{1,N-3} = F_{1,N-2} \cdot \frac{N-3}{N-2} \cdot \frac{\left[\dfrac{s^2_{Y_T}\ (1-r^2_{XY_T})}{s^2_{Y_W}\ (1-r^2_{XY_W})} - 1\right]}{\left[\dfrac{s^2_{Y_T} + s^2_{X_T} - 2r_{XY_T}\, s_{Y_T}\, s_{X_T}}{s^2_{Y_W} + s^2_{X_W} - 2r_{XY_W}\, s_{Y_W}\, s_{X_W}} - 1\right]} \qquad (2)$$

__Proof__ $F_{1,N-2} = (N-2)\left[\dfrac{TSSG}{WSSG} - 1\right]$, where TSSG and WSSG are the total and within sums of squares for gain scores, respectively

$$= (N-2)\left[\frac{Ns_{G_T}^2}{Ns_{G_W}^2} - 1\right],$$ where $s^2_{G_T}$ and $s^2_{G_W}$ are the variances of the gain scores for total and within

$$= (N-2)\left[\frac{s^2_{(Y-X)_T}}{s^2_{(Y-X)_W}} - 1\right]$$

$$= (N-2)\left[\frac{s^2_{Y_T} + s^2_{X_T} - 2r_{XY_T}\, s_{Y_T}\, s_{X_T}}{s^2_{Y_W} + s^2_{X_W} - 2r_{XY_W}\, s_{Y_W}\, s_{X_W}} - 1\right]$$

From Theorem 1:

$$F_{1,N-3} = (N-3)\left[\frac{s^2_{Y_T}(1-r^2_{XY_T})}{s^2_{Y_W}(1-r^2_{XY_W})} - 1\right]$$

$$\frac{F_{1,N-3}}{F_{1,N-2}} = \frac{(N-3)\left[\dfrac{s^2_{Y_T}(1-r^2_{XY_T})}{s^2_{Y_W}(1-r^2_{XY_W})} - 1\right]}{(N-2)\left[\dfrac{s^2_{Y_T} + s^2_{X_T} - 2r_{XY_T}\, s_{Y_T}\, s_{X_T}}{s^2_{Y_W} + s^2_{X_W} - 2r_{XY_W}\, s_{Y_W}\, s_{X_W}} - 1\right]}$$

$$F_{1,N-3} = \frac{F_{1,N-2}\cdot\dfrac{N-3}{N-2}\cdot\left[\dfrac{s^2_{Y_T}(1-r^2_{XY_T})}{s^2_{Y_W}(1-r^2_{XY_W})} - 1\right]}{\left[\dfrac{s^2_{Y_T} + s^2_{X_T} - 2r_{XY_T}\, s_{Y_T}\, s_{X_T}}{s^2_{Y_W} + s^2_{X_W} - 2r_{XY_W}\, s_{Y_W}\, s_{X_W}} - 1\right]}$$

REFERENCES

Campbell, Donald S., and Stanley, Julian C. "Experimental and Quasi-Experimental Designs for Research on Teaching". In Gage, N.L. (Ed.) Handbook of Research on Teaching. Chicago: Rand McNally and Co., 1963. 1218 pp.

Edwards, Allen L. Experimental Design in Psychological Research (third edition). New York: Holt, Rinehart, and Winston, Inc., 1968. 455 pp.

Feldt, Leonard S. "A Comparison of the Precision of Three Experimental Designs Employing a Concomitant Variable". Psychometrika, 23: 335-353; December, 1958.

McNemar, Quinn. Psychological Statistics (third edition). New York: John Wiley and Sons, Inc., 1962. 451 pp.

Olkin, Ingram. "Correlations Revisited". In Julian C. Stanley (Ed.) Improving Experimental Design and Statistical Analysis. Seventh Annual Phi Delta Kappa Symposium on Educational Research. Chicago: Rand McNally and Company, 1967. 305 pp.

Wilson, James W., and Carry, L. Ray. "Homogeneity of Regression - Its Rationale, Computation and Use". American Educational Research Journal, 6: 80-90; January, 1969.